Catalogue no. 13-604-M
ISSN 1707-1739
ISBN 978-1-100-25571-2

Income and Expenditure Accounts Technical Series

Constructing Provincial Time Series: A Discussion of Data Sources and Methods

by Ryan Macdonald

Release date: February 12, 2015

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at infostats@statcan.gc.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "About us" > "The agency" > "Providing services to Canadians."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p} preliminary
^{r} revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E} use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2015

All rights reserved. Use of this publication is governed by the Statistics Canada Open Licence Agreement.

An HTML version **is also available.**

Cette publication est aussi disponible en français.

Table of contents

1 Introduction **4**

2 Principles **4**

3 Linking methodology **5**

4 Linking procedures by variable **6**

4.1 Household income, household disposable income and compensation of employees 6

5 City-based consumer price indexes **6**

6 Population **7**

References **9**

Acknowledgements

The author would like to thank Mark Brown for his insight into the behavior of regional economies, John Baldwin for his guidance, and James Tebrake and Karen Wilson for their comments, suggestions and recommendations for working with historical data from the Canadian System of National Accounts.

1 Introduction

This paper describes the linking and estimation procedures used to generate an updated vintage of the provincial comparison dataset for the period from 1926 to 2011 employed in Brown and Macdonald (2015) to examine convergence across Canadian provinces. The dataset was expanded to include data from 2012 and the Yukon, the Northwest Territories, and Nunavut.

The long-run data discussed here start in 1926 and cover variables related to household income, population and consumer prices.

This new dataset increases the information available for comparing the performance of provinces and territories across a range of measures. It combines often fragmented provincial time series data that, as such, are of limited utility for examining the evolution of provincial economies over extended periods. More advanced statistical methods, and models with greater breadth and depth, are difficult to apply to existing fragmented Canadian data. The longitudinal nature of the new provincial dataset remedies this shortcoming.

This report explains the construction of the latest vintage of the dataset. The dataset contains the most up-to-date information available. It is not identical to that employed in Brown and Macdonald (2015), as it incorporates improved linking procedures, and revised data sources.

This document is structured as follows. Section 2 describes the principles applied to the dataset. Section 3 explains the rationale underlying the linking procedures. Section 4 then discusses how the historical time series for each variable was constructed.

2 Principles

The new provincial dataset is designed to meet four criteria: quality, relevance, expandability, and ability to be updated. The four criteria are discussed in order of importance.

The first criteria is quality. To meet this goal, official sources are used whenever possible, notably, estimates from Statistics Canada, from the Dominion Bureau of Statistics, and from the Historical Statistics of Canada. If official estimates are not available, sources such as academic studies are used to improve coverage and time span. In all cases, the data source and how it was used are documented to allow for replication and verification.

The second criteria is relevance. To accomplish this, modern estimates from the Canadian System of National Accounts are used, and the data are linked in such a way that they can be updated as additional data points become available. Consequently, all variables are produced so as to link historical data to modern data. Specifically, historical estimates are adjusted to produce back-cast estimates that do not exhibit structural breaks or level shifts when changes are made in the vintage of data employed.

Third, the dataset must be expandable—over time and across the indicators. All expansions that result in new data points going forward through time need to be made in a timely fashion. All expansions that lead to new variables being added to the dataset, or that increase the span of the data historically need to be published with documentation.

Fourth, the dataset should be open to revision. Because it is a research dataset, knowledge acquired through its use can be applied to improve quality. For instance, better data sources may be located, and if the dataset is expanded, the need for consistent estimates across a range of variables may necessitate new and different linking procedures. All revisions should be released with documentation.

3 Linking methodology

Data sources are ranked from newest to oldest (for example, Statistics Canada, Dominion Bureau of Statistics, academic sources, etc.), and the data are then linked using the most appropriate method.

The starting point for each variable is the current vintage of data. The current vintage corresponds with the price index, income and population variables produced on an ongoing basis by Statistics Canada. In the dataset, they are denoted by a superscript (A) to indicate their current vintage status.

When national accounts programs are revised, recent data tend to be most affected, and historical estimates, least affected. This is especially the case for growth rates. The relatively larger adjustment to recent data occurs as statistical systems react to structural changes in the economies about which they report.

In most cases, historical estimates are linked based on their growth rates; that is, the modern estimates are projected back through time based on the growth rates of historical estimates for similar variables. In the dataset, these estimates are denoted by a superscript (B) to indicate that they are historical projections. The historical estimates are treated as if they are indexes with a base value equal to the earliest data point from the modern data vintage. This approach works well because, while revisions may change level estimates, the growth rates for the economic aggregates employed are similar across data vintages.

The advantage of using growth rates rather than regressions to back-cast the time series is that the magnitude of historical events is preserved. Regression estimates can expand or contract the variance of historical estimates relative to current estimates based on what is typically a limited sample of overlapping observations. Moreover, the overlapping observations may not be particularly representative of previous periods.

In some cases, data are limited, and instrumental variable techniques (nearest-neighbour matches, regression-based approaches, constructed indicators) are used. These estimates are denoted by a superscript (C) to indicate the instrumental variable approach.

4 Linking procedures by variable

4.1 Household income, household disposable income and compensation of employees

Estimates of household income, household disposable income and compensation of employees are drawn from two sources. Modern estimates for household income and household disposable income are from CANSIM table 384-0040. Household income estimates are composed of primary household income and current transfers received. Household disposable income equals personal income less current transfers paid. Compensation of employees is taken from CANSIM table 384-0037 and is equal to wages and salaries paid in Canada, plus employer social contributions.

The historical estimates are taken from the National Income and Expenditure Accounts Annual Estimates from 1926 to 1986. Household income is taken from Table 35, and corresponds to what is currently referred to as primary income plus current transfers. Household disposable income is taken from Table 37. A projector for compensation of employees is taken from Table 38, and it corresponds to wages, salaries, and supplementary labour income.

For the modern and historical vintages employed here, the units are not identical, but are reasonably comparable through time. The household sector covers persons and unincorporated businesses in both vintages. However, differences in imputation methods for items such as income from owner-occupied housing, or for compositional shifts at the margin for entries such as clergy income, affect the absolute levels. The largest conceptual difference is for compensation of employees, which includes a balance of payments adjustment after 1981.

For all provinces, the growth rates of the historical series are used to back-cast estimates of provincial levels and for outside Canada. For the territories, the data are less detailed. In historical estimates from 1926 to 1950, the territories are included with British Columbia. From 1951 to 1986, the historical estimates combine Yukon and an aggregate of the Northwest Territories that includes Nunavut. The data for the territories are, therefore, combined into a territorial aggregate that is used to back-cast estimates in earlier periods. As more detailed data became available at later dates, values for Yukon and the Northwest Territories including Nunavut are presented. Finally, starting in 1999, separate series for the Northwest Territories and Nunavut are presented. Estimates for Newfoundland and Labrador begin in 1949.

5 City-based consumer price indexes

Historical consumer price indexes (CPIs) are not available by province. However, they are available for major cities. This leaves two options for creating historical CPI estimates. One is to link the current vintage provincial CPI to historical estimates of city-specific CPIs. The second is to link city CPIs through time, and apply them to provincial income estimates to produce real income series. The latter method was chosen because it yields more consistency in the price series over time, which facilitates examinations beyond those in Brown and Macdonald (2015). It also results in a more consistent measurement error over time, which provides more stability in the direction of biases.

Table 1
City consumer price indexes, sample spans for measured data and instruments

City	Province	Measured sample span	Instruments
St. John's	Newfoundland and Labrador	1952 to present	Average of Halifax, Nova Scotia, and Saint John, New Brunswick, for 1949 to 1951
Charlottetown and Summerside	Prince Edward Island	1974 to present	Saint John for 1926 to 1973
Halifax	Nova Scotia	1926 to present	...
Saint John	New Brunswick	1926 to present	...
Montréal	Quebec	1926 to present	...
Toronto	Ontario	1926 to present	...
Winnipeg	Manitoba	1926 to present	...
Regina	Saskatchewan	1926 to present	...
Calgary	Alberta	1926 to present	...
Vancouver	British Columbia	1926 to present	...
Whitehorse	Yukon	1983 to present	...
Yellowknife	Northwest Territories	1983 to present	...
Iqaluit	Nunavut	2003 to present	...

... not applicable
Source: Statistics Canada.

Price indexes for a major city from each province from three sources are linked. These indexes provide long-term measures of price growth. For 1971 to 2011, the data are from CANSIM table 326-0021. For 1940 to 1970, city-CPI estimates are from Series K23 of the Historical Statistics of Canada. For 1926 to 1939, the data source is Emery and Levitt (2002). These sources do not provide as much breadth through time as income or population estimates (Table 1). Data for St. John's, Newfoundland and Labrador, begin in 1952. An average of price movements from Halifax, Nova Scotia, and Saint John, New Brunswick, are used to back-cast St. John's, Newfoundland and Labrador, for 1949 to 1952. Estimates for Charlottetown and Summerside, Prince Edward Island, begin in 1974. To produce historical estimates, the series for Saint John, New Brunswick, is used to back-cast the historical series.

The use of cities in neighbouring provinces when data are missing provides a balanced panel dataset. This facilitates internal instrumental variable estimation, but it also introduces measurement error. The trade-off between including or omitting these data points depends on the use of the data and the methods chosen for estimation.

6 Population

Population data are drawn directly from CANSIM tables 380-0043 and 051-0005. The historical data in table 380-0043 are based on an annual series, where population is measured on June 1. The modern data in table 051-0005 are based on quarterly observations. They are converted to an annual frequency by selecting the Q2 stocks from each year to approximate the mid-year sampling of the historical data.

Despite a minor level difference between the modern and historical series, the levels of the two series were linked "as is" to avoid creating a second set of population estimates for the historical period. When rounded to the nearest thousand, the difference between the two data sources is minor.

The historical file is used for data points between 1926 and 1951. The modern series is used from 1952 to the present. The aggregations for the provinces link well. For the territories, an aggregate of Yukon and the Northwest Territories including Nunavut is reported for the 1926-to-present period. Yukon and the Northwest Territories including Nunavut are also reported separately over this period. Beginning in 1999, the Northwest Territories and Nunavut are reported individually. Estimates for Newfoundland and Labrador begin in 1949.

References

Brown M., and R. Macdonald. 2015. *Provincial Convergence and Divergence in Canada, 1926 to 2011.* Economic Analysis Research Paper Series, no. 096. Statistics Canada Catalogue no. 11F0027M. Ottawa: Statistics Canada.

Emery, J.C., and C. Levitt. 2002. "Cost of living, real wages and real incomes in thirteen Canadian cities 1900 to 1950." *Canadian Journal of Economics* 35 (1): 115–137.